the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2020) doi: 10.1111/tpj.14962

Identification of transcription factors controlling floral morphology in wild *Petunia* species with contrasting pollination syndromes

Tural Yarahmadov[1,2,†], Sarah Robinson[1,3,†], Mathieu Hanemian[1,4,†], Valentin Pulver[1] and Cris Kuhlemeier[1,*]

[1]*Institute of Plant Sciences, University of Bern, Altenbergrain 21, Bern CH-3013, Switzerland,*

[2]*Department of BioMedical Research, University of Bern, Bern CH-3008, Switzerland,*

[3]*Sainsbury Laboratory, University of Cambridge, Cambridge CB2 1LR, UK, and*

[4]*LIPM, Université de Toulouse, INRAE, CNRS, Castanet-Tolosan, France*

Received 4 October 2019; accepted 15 July 2020.

*For correspondence (e-mail cris.kuhlemeier@ips.unibe.ch).

†These authors contributed equally to this work.

SUMMARY

Adaptation to different pollinators is an important driver of speciation in the angiosperms. Genetic approaches such as QTL mapping have been successfully used to identify the underlying speciation genes. However, these methods are often limited by widespread suppression of recombination due to divergence between species. While the mutations that caused the interspecific differences in floral color and scent have been elucidated in a variety of plant genera, the genes that are responsible for morphological differences remain mostly unknown. Differences in floral organ length determine the pollination efficiency of hawkmoths and hummingbirds, and therefore the genes that control these differences are potential speciation genes. Identifying such genes is challenging, especially in non-model species and when studying complex traits for which little prior genetic and biochemical knowledge is available. Here we combine transcriptomics with detailed growth analysis to identify candidate transcription factors underlying interspecific variation in the styles of *Petunia* flowers. Starting from a set of 2284 genes, stepwise filtering for expression in styles, differential expression between species, correlation with growth-related traits, allele-specific expression in interspecific hybrids, and/or high-impact polymorphisms resulted in a set of 43 candidate speciation genes. Validation by virus-induced gene silencing identified two MYB transcription factors, EOBI and EOBII, that were previously shown to regulate floral scent emission, a trait associated with pollination by hawkmoths.

Keywords: speciation genes, *Petunia*, transcription factors, MYB, style, floral morphology, transcriptomics, growth analysis, biomechanics, ACME, RNAseq.

INTRODUCTION

A key innovation in the angiosperms is the recruitment of animal pollinators to enhance reproductive success (Stebbins, 1970; Thomson and Wilson, 2008; Armbruster and Muchhala, 2009). The remarkable differences in floral color, scent, and morphology, even between closely related species, are thought to have evolved in response to selective pressure imposed by corresponding classes of pollinators to form so-called pollination syndromes (Fenster *et al.*, 2004). Visitation by different pollinators is likely to promote reproductive isolation and eventually lead to speciation. Consequently, the genes that caused the differences between these floral pollination syndromes can be classified as 'speciation genes' and the identification of such genes is a major goal of evolutionary genetics.

Substantial progress has been made regarding the genes underlying floral attraction traits, such as color and scent, for example (Hoballah *et al.*, 2007; Wessinger and Rausher, 2012; Amrad *et al.*, 2016; Peng *et al.*, 2017), but far less is known about the molecular-genetic basis of differences in floral morphology. The specific morphology of a flower serves to attract specific guilds of pollinators and ensure efficient pollination (Alexandersson and Johnson, 2002; Fenster *et al.*, 2004; Cronk and Ojeda, 2008). Bees tend to visit short and wide flowers that provide a landing platform for the animal (Whitney *et al.*, 2009), whereas hawkmoths and hummingbirds hover in front of the flower

while feeding on the nectar (Raguso and Willis, 2002; Sapir and Dudley, 2013). The long and narrow corolla tubes limit access to bees and other visitors. Further elongation of all the floral organs will restrict effective pollination to hawkmoths with a longer proboscis, which is expected to increase specificity but may compromise reproductive assurance (Alexandersson and Johnson, 2002). Differential exsertion of the reproductive organs is likely to reduce selfing and may also optimize pollen transfer and/or deter visitation by hawkmoths (Alexandersson and Johnson, 2002; Fenster *et al.*, 2004; Cronk and Ojeda, 2008).

To study the genetic basis of the differences in reproductive organ morphology, we used three closely related species in the South American genus *Petunia* (Figure 1a). The hawkmoth-pollinated *Petunia axillaris* and *Petunia parodii* have a white, UV-absorbing corolla, emit a complex blend of phenylpropanoid volatiles, and have a slightly inserted style. Tube, style, and stamens are significantly longer in *P. parodii* (7 cm versus 4 cm in *P. axillaris*), which may attract hawkmoths with a long enough proboscis, increasing the specificity of pollen transfer but also reducing reproductive assurance (Venail *et al.*, 2010). The hummingbird-pollinated *Petunia exserta* has a deep red, UV-reflective corolla, with reflexed lobes, and is odorless. The style protrudes above the rim of the corolla and most often also above the anthers.

QTL analysis has identified three or four loci responsible for the difference in *Petunia* style length (Venail *et al.*, 2010; Hermann *et al.*, 2015). In a *P. axillaris* x *P. parodii* F_2 population, there was co-segregation of tube, style, and stamen length, suggesting that the genetic loci involved have a general effect on floral elongation. Crosses between *P. axillaris* and *P. exserta* indicate potentially independent effects on style and stamen length. However, these results should be interpreted with caution, as the QTL intervals are quite large and might fragment upon further fine mapping. For instance, a style length QTL is located within a genomic region of severely suppressed recombination on chromosome II (Hermann *et al.*, 2013) and it is conceivable that such a QTL contains multiple genes with separate effects on individual traits. Suppression of recombination appears to be a common feature of interspecific crosses (Jia *et al.*, 2012; Ostberg *et al.*, 2013; Ren *et al.*, 2018). This makes further fine mapping problematic and therefore, alternative methods for identifying speciation genes are clearly needed.

As style length can be easily measured and shows little environmental variation, we focused our research on this character. Mechanistically, the final length of the style depends on cell wall properties and cell division rate, as well as growth rate and duration, with each of these processes being potential targets of speciation (Lockhart, 1965; Beemster and Baskin, 1998; Czesnick and Lenhard, 2015; Peaucelle *et al.*, 2015; Braybrook and Jönsson, 2016). Therefore, we decided to perform a detailed analysis of style growth and searched for correlations between growth properties and gene expression profiles at different developmental time points. We focused our analysis on transcription factors on the assumption that these are overrepresented among the genes underlying the evolution of plant development (Doebley and Lukens, 1998) and this may apply for genes underlying shifts in pollination syndromes as well. It has even been suggested that one class of transcription factors, the MYBs, may be preferred targets of pollination syndrome evolution, analogous to the predominance of MADS box transcription factors in specifying floral whorl identity (Yuan *et al.*, 2013).

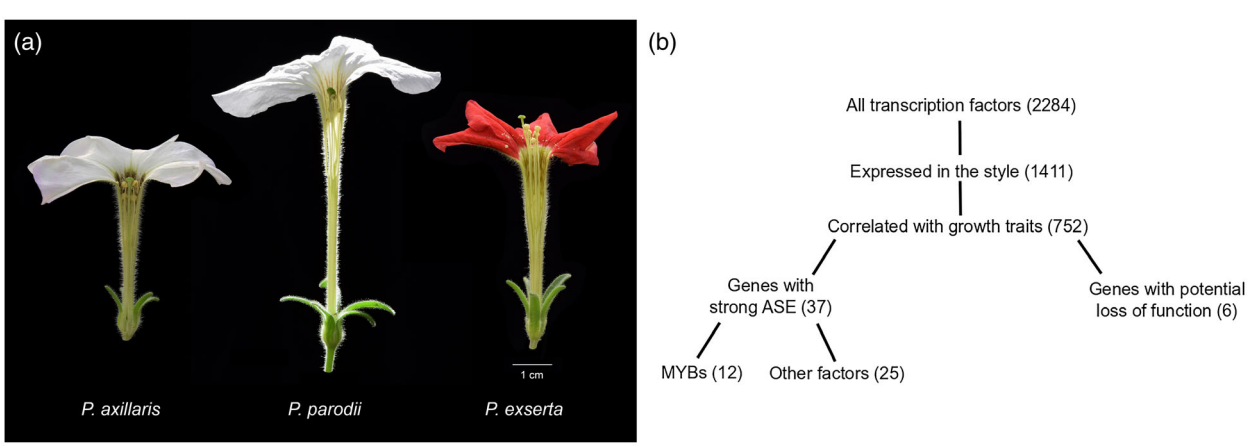

Figure 1. Workflow for the identification of candidate speciation genes in *Petunia* species with differences in reproductive organ morphology.
(a) *P. axillaris*, *P. parodii*, and *P. exserta* with corolla partly removed to show the reproductive organs.
(b) The initial dataset is filtered stepwise based on expression in styles and correlation with the growth traits of interest. The reduced dataset is then analyzed for differential and allele-specific expression (ASE) as well as coding sequence differences between the studied species. The numbers in the brackets represent the number of genes remaining after each step.

Here we present a workflow for the identification of candidate speciation genes affecting *Petunia* style growth (Figure 1b). Starting from the set of all *Petunia* transcription factors we applied different filters: (i) expression in the style in at least one species; (ii) correlation of transcript levels with growth-associated traits; (iii) allele-specific expression (ASE) in interspecific F_1 plants; and (iv) functional variants in the coding region. This hierarchical filtering led to a set of 43 candidate genes enriched for MYB transcription factors (14 genes). Virus-induced gene silencing (VIGS) provided functional validation of two MYB transcription factors that were previously shown to regulate floral volatile emission in *P. hybrida* (Spitzer-Rimon *et al.*, 2010; Van Moerkercke *et al.*, 2011; Spitzer-Rimon *et al.*, 2012).

RESULTS

Petunia floral organs of different species have different growth-associated trait profiles

To understand how *Petunia* floral reproductive organs attain different final lengths, measurements of the growing buds were taken from less than 1 cm long until 1–2 days after flower opening. As flower opening is a reliable and easily measurable time point, all data were expressed as hours until opening. The length of the buds at the time of opening is significantly different between the three species (Figure 2a). The growth curves of the buds of all species show a characteristic sigmoidal type growth: with initial slow growth followed by fast elongation and then slowing again around the time of opening (Figure 2b).

As the pistil develops inside the bud, we dissected buds at different stages of growth and determined the relationship between bud length and pistil length (Figure 2c). Pistil length correlates linearly with bud length, and therefore, bud growth can be used as a proxy for pistil growth. The early growth is highest in *P. exserta,* while *P. parodii* elongates the fastest during the steep part of the exponential growth curve (Figure 2d).

To determine what might cause the difference in style length between the three *Petunia* species, we analyzed key growth parameters, namely growth rate, cell wall elasticity, and cell division rate (Figure 3). We defined four developmental stages: the small stage (S) is the early stage when the buds are less than 1 cm in length; the medium-1 stage (M1) is when the pistils are growing rapidly; the medium-2 stage (M2) is when the growth rate is most different between the three species; and the long stage (L) is when the flowers have been open for less than 1 day. The growth rate is highest in the M1 stage (Figure 3a). Only *P. parodii*, which has the highest growth rate, maintains a similar level in the M2 stage, while the growth rate of *P. axillaris* and *P. exserta* decreases after the M1 stage. We approximated the elasticity of the cell walls of the styles by measuring the relative change in length, known as strain, when the same amount of force was applied to them (see Experimental procedures). In all species the cell walls were more elastic at the S stage, and then became stiffer with developmental time (Figure 3b), until the L stage, when they could not be extended at all. More specifically, *P. exserta* styles exhibited significantly more strain during the mechanical tests, suggesting they are more elastic than the two other species in the M1 stage. The measured strain

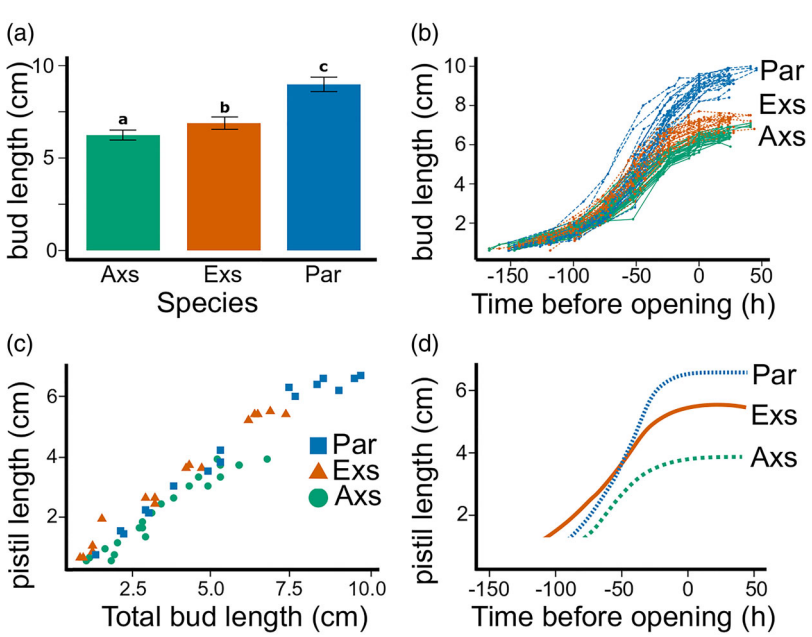

Figure 2. Bud and pistil growth in *P. axillaris* (Axs), *P. exserta* (Exs), and *P. parodii* (Par).
(a) Total length of buds at the time of opening. Bars show means ± SD. *P. axillaris*, mean = 6.2 ± 0.27 cm, $n = 31$ buds; *P. exserta*, mean = 6.9 ± 0.33 cm, $n = 30$ buds; *P. parodii*, mean = 9.0 ± 0.39 cm, $n = 27$ buds. Bars with different letters are significantly different ($P < 2.9\ e{-}11$, Welch *t*-test).
(b) Bud growth curves on the basis of time to flower opening.
(c) Relationship between bud and pistil length (*P. axillaris*, $n = 23$; *P. exserta*, $n = 16$; *P. parodii*, $n = 16$).
(d) The data on bud length against time to flower opening (b) were combined with the data on pistil length against bud length (c) to produce an estimate of pistil length against time. Time is expressed as hours before opening.

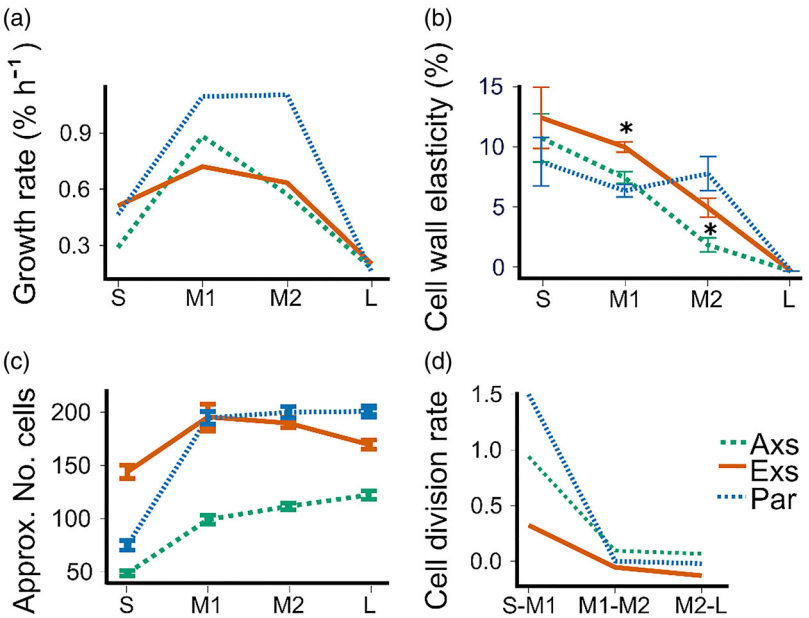

Figure 3. Growth rate, wall elasticity (strain), and cell division rate of the styles change throughout development.
(a) The relative growth rate of the pistils was extracted for each species at each stage from the growth curves shown in Figure 2(d).
(b) The mechanical properties of the styles were analyzed by applying an oscillating force and measuring the strain. The strain is significantly different between *P. exserta* and the other two species at the M1 stage ($n \geq 4$ per species, $P < 0.05$) and between *P. axillaris* and the other two species at the M2 stage ($n \geq 4$ per species, $P < 0.05$).
(c) Cell number in the different species. The number of cells in the styles at the different stages was estimated. Four style sections from each species from each stage were imaged. At least 10 cell lengths were measured. The average cell number was calculated by dividing the length of the style by the average cell length.
(d) The relative change in cell number from one stage to the next was computed to obtain the cell division rate. Note: the method resulted in a slight reduction in the number of cells in *P. exserta* in the L stage versus M2. Since there is no evidence of cell death or cells sliding relative to each other, we assume the cell division rate to be zero. Error bars show the SD.

was significantly lower for *P. axillaris* in the M2 stage, suggesting it is less elastic than the other samples at this stage. The number of cells increases with time, with the most division occurring between the S and the M1 stages in all the species (Figure 3c,d). *P. parodii* and *P. exserta* show the highest and the lowest division rates at this early stage of style development, respectively. However, this is likely due to the high number of cells already present in the S stage of *P. exserta,* suggesting that cell division occurred even earlier in this species. Overall, the three species show similar trends in their development but there are also some notable differences.

A workflow for speciation gene identification

Figure 1(b) summarizes the workflow to identify potential speciation genes, starting from the complete set of 2284 *Petunia* genes encoding transcription factors. These putative transcription factors were identified by comparing publicly available *Petunia* protein sequences with domain signatures of known plant transcription factors. BLAST searches were also performed to identify *Arabidopsis* homologs in *Petunia* and vice versa. The steps of this workflow are sequentially described in details below.

1) Expression in the style in at least one species. We conducted an RNAseq experiment to determine which of these putative *Petunia* transcription factors were expressed in the style. RNA was extracted from styles of the three parental species at the four developmental stages described above. Clustering analysis of gene expression showed that samples are more similar according to the developmental stage than the species (Figure S1). The MS and SS expression vectors each have homogeneous clusters which are more closely related to each other than they are to either the MS2 or the LS stage. The clusters of the MS2 and LS samples are not homogenous and cluster together. We also observed that the transcriptome of the S stage stands apart from the other stages consistent with what was observed for growth rate, cell number, and cell division rate. The genes were filtered to remove pollen genes that contaminated the late-stage style samples (see Experimental procedures for details). Genes showing no or very low expression were removed using the R package HTSfilter (Rau *et al.*, 2013). From 2284 putative *Petunia* transcription factors initially identified, 1411 were expressed in styles according to these criteria.

2) Correlation of transcript levels with growth-associated traits. Our analysis showed that *Petunia* styles grow by a mixture of cell division in early stages and cell elongation in the later stages. To identify genes that might regulate one of the growth-related traits, we performed a correlation analysis between each normalized gene expression profile and total pistil length, cell wall elasticity, cell division rate, and growth rate. The expression of genes across all species and all stages was correlated with the growth-associated trait values. Considering a Pearson's correlation coefficient (PCC) $\geq$ 0.75 by the absolute value (further referred to as strongly correlating), the expression of 752 genes correlated with at least one of the four measured traits.

3) Allele-specific expression (ASE) in interspecific F_1s. Differences in gene expression may be caused in cis, that is, by differences within the gene and its regulatory sequences, or in trans, that is, by the differential activity of a non-linked transcription factor. To check whether the observed differential expression is due to differences in cis or in trans, an analysis of ASE in interspecific F_1 hybrids was performed (Pastinen, 2010; Wittkopp and Kalay, 2012). RNAseq was carried out using RNA from the F_1 hybrids between each of the wild-type parents at growth stages M1 to L. Out of 752 genes strongly correlating with growth-associated traits, 38 displayed strong ASE in at least one interspecific F_1 hybrid (ASE > 0.75). At this stage, sequences of these 38 genes and the annotations they are based on were manually inspected. After the inspection, one of them was discarded as likely not encoding a transcription factor, creating a set of 37 genes of interest (Table 1).

4) Functional variants in the coding region. In addition to the differences detected by ASE we evaluated the impact of potential divergence in protein function. The coding sequences of the 752 transcription factors whose expression correlated with growth-associated traits were compared between the *Petunia* species. The RNAseq readmaps were processed to detect and predict the impact of SNPs. SNPs from the 'high-impact' category (e.g., frameshifts, loss and gain of start/stop codons, rare amino acid variants, and exon splice site mutations) were considered for further analysis. To further validate the existence of target genes and SNPs, comparison of genomic sequences of genes of interest from different species was performed using BLAST+. However, whole genome sequences were available only for *P. axillaris* and *P. exserta*. Therefore, we assembled a de novo draft genome for *P. parodii* using high-coverage short-read sequencing (Table 2). In total, we found six genes with amino acid variation of potential functional significance using homologous sequences between *Petunia* species (Table 3). This brings the total dataset of candidate speciation genes to 43 (Tables 1, S2, and S3).

Analysis of the potential speciation genes

Of the 43 genes identified, 32 genes had expression levels strongly correlated with either cell wall properties or cell division (Figure 4). Most genes with expression levels that correlated with cell wall elasticity also correlated with style length (12/17). This reflects the strong correlation between cell wall properties and total length that accompanies the progression through the developmental stages (see also Figures 3b and 2d). The majority of genes correlating with style length show a positive correlation with the trait, while the opposite is observed for cell wall elasticity (Figure 5). Comparably few genes (5/43) displayed expression profiles that correlated with the growth rate. This suggests that the observed growth rate is more likely a result of cell elongation and cell division and that these traits are regulated independently by different sets of genes.

Interestingly, MYB transcription factors were overrepresented with 14 genes out of 43 belonging to this family, which is highly unlikely by chance (Table S4). An additional point of interest is that three of these MYBs are known regulators of the phenylpropanoid pathway in *Petunia*: scent-related factors ODO1, EOBI, and PhMYB4 (Colquhoun *et al.*, 2011a); (Van Moerkercke *et al.*, 2011); (Spitzer-Rimon *et al.*, 2012). Moreover, all three genes show a strong positive correlation with style length and a negative correlation with cell wall elasticity. Such a correlation is seen in nine out of the 14 MYB genes. We also noticed *Petunia* genes similar to *Arabidopsis* genes known to affect plant organ morphology, such as *ARF2*, *SHR*, and *BPE* in the final dataset (Tables 1 and S2).

Candidate gene validation

To determine whether the final list of candidate genes identified have a role in controlling floral morphology, functional validation was performed using VIGS for a subset of genes based on an additional literature review. The homeobox gene *ATHB13* is involved in the elongation of inflorescence stems and the HLH gene *BIGPETAL* regulates petal size in *Arabidopsis thaliana* (Ribone *et al.*, 2015); (Szécsi *et al.*, 2006). For the MYB genes, we selected *PhMYB4*, *EOBI*, and *ODO1* together with *EOBII*, which were shown to regulate scent production in *P. hybrida* (Verdonk *et al.*, 2005; Spitzer-Rimon *et al.*, 2010; Colquhoun *et al.*, 2011a; Spitzer-Rimon *et al.*, 2012). *ODO1*/*EOBI*/*EOBII* and their homologs have pleiotropic effects on flower development in *Petunia, Nicotiana,* and *Arabidopsis* (Colquhoun *et al.*, 2011b; Liu and Thornburg, 2012; Battat *et al.*, 2019). *EOBII*, a close homolog of *EOBI*, does not appear in Table 1 because the lack of SNPs between the transcripts precluded ASE analysis. Because of its relationships with *EOBI* and *ODO1*, we added it to our VIGS analysis. The expression patterns of the selected genes

Table 1 Growth-related transcription factors with allele-specific expression

Gene ID	Correlated trait	Predicted TF family	Most similar to *Arabidopsis*
Peaxi162Scf00005g00506	L (−), CWE (+)	AP2	*TOE1*
Peaxi162Scf00074g00101	CDR (−)	AP2	
Peaxi162Scf00332g00746	CDR (−)	AP2	*SHN1*
Peaxi162Scf01168g00016	CWE (+)	AUX/IAA	*ARF2*
Peaxi162Scf00418g00083	L (+), CWE (−)	bZIP	*BZIP11*
Peaxi162Scf01290g00246	CDR (+)	bZIP	
Peaxi162Scf00017g02335	CWE (+)	E2F/DP	*E2FE*
Peaxi162Scf00548g00063	L (−), CWE (+)	GRAS	*SHR*
Peaxi162Scf00209g00932	GR (+), CDR (−)	bHLH	*BPE*
Peaxi162Scf00827g00055	CDR (−)	Histone-like (CBF/NF-Y)	*L1L*
Peaxi162Scf00026g00296	GR (−)	Homeobox	*ATHB13*
Peaxi162Scf00129g00319	L (+), CDR (−)	Homeobox	*HB3*
Peaxi162Scf00740g00325	L (+), CWE (−)	LIM	
Peaxi162Scf00002g00037 (*ODO1*)	L (+), CWE (−)	MYB	*MYB42*
Peaxi162Scf00016g00729	L (+)	MYB	*RL6*
Peaxi162Scf00042g02519	L (+), CWE (−)	MYB	*MYB4*
Peaxi162Scf00048g01417	CDR (+)	MYB	*RVE1*
Peaxi162Scf00064g00423	GR (+)	MYB	*MEE3*
Peaxi162Scf00102g01226 (ODO1-like)	L (+), CWE (−)	MYB	*MYB42*
Peaxi162Scf00129g01231 (EOBI)	L (+), CWE (−)	MYB	*MYB21*
Peaxi162Scf00185g01625	CWE (−)	MYB	
Peaxi162Scf00266g00512	L (+)	MYB	*MYBD*
Peaxi162Scf00366g00316	CDR (−)	MYB	*RL6*
Peaxi162Scf00401g00025	L (+), CWE (−)	MYB	*DIV1*
Peaxi162Scf01221g00042 (PhMYB4)	L (+), CWE (−)	MYB	*MYB4*
Peaxi162Scf00074g00543	CDR (+)	MYC/MYB (NIN-like)	NLP7
Peaxi162Scf00045g00127	L (+), CWE (−)	PLATZ	
Peaxi162Scf00031g01525	L (−), CWE (+)	SBP	SPL3
Peaxi162Scf00656g00007	L (−)	SHI-like	STY1
Peaxi162Scf00013g00725	CDR (+)	SRF-type (MADS)	APETALA1
Peaxi162Scf00483g00315	L (+)	TAZ zinc finger	BT3
Peaxi162Scf00317g00117	GR (+), CDR (−)	tify	JAZ8
Peaxi162Scf00332g00071	L (+), CDR (−)	Tubby C2	
Peaxi162Scf00007g00315	L (+), CWE (−)	WRKY	WRKY6
Peaxi162Scf00459g00841	CDR (+)	WRKY	WRKY11
Peaxi162Scf00016g00288	L (+), CWE (−)	Zinc finger, C2H2 type	ZAT10
Peaxi162Scf00029g02912	CDR (−)	Zinc finger, C2H2 type	WIP2

This table lists characteristics of the transcription factors (TFs) correlated with at least one growth-related trait and displaying allele-specific expression. In the 'correlated trait' column, (+) and (−) mean that a given trait is correlated or anti-correlated with the expression of a given gene. L, pistil length; CWE, cell wall elasticity; CDR, cell division rate; GR, pistil growth rate. For details of the analysis, see Tables S2 and S3.

are shown in Figure 6(a). Based on stringent criteria for significance (see Experimental procedures), pistil length was reduced by the VIGS construct targeting *EOBI* in *P. axillaris,* and by the VIGS construct targeting *EOBII* in both *P. axillaris* and *P. exserta* (Figure 6b).

DISCUSSION

Wild *Petunia* species are adapted to pollination by bees, hawkmoths, or hummingbirds and thus offer the opportunity to study pollinator-driven speciation. Our interest is the molecular basis of pollinator-mediated speciation, which presents some unique challenges. First, genetic and

Table 2 Summary of the short-read *P. parodii* reference genome assembly

Total size	0.9 Gb
Average read coverage	49×
Mean insert size (nt)	523
Contig number	3006811
Mean contig size (nt)	314
Median contig size (nt)	147
Longest sequence (nt)	31531
N50 (nt)	526

N50 refers to the length of the shortest contig at 50% of the total genome length. Genome statistics were estimated using samtools v.1.6 (Li *et al.*, 2009).

Table 3 Growth-related transcription factors with potential loss of function-inducing polymorphisms between studied species

Gene ID	Predicted mutation	Predicted TF family	*Arabidopsis* homolog
Peaxi162Scf00071g00035	Loss of stop codon (Exs)	bHLH	*AIF4*
Peaxi162Scf00171g00419	Frameshift (Exs)	MYB	*RVE1*
Peaxi162Scf00330g00619	Frameshift (Par)	Homeobox	
Peaxi162Scf00390g00336	Frameshift (Axs)	MYB	*TRFL5*
Peaxi162Scf00548g00014	Frameshift (Exs)	bZIP	*bZIP6*
Peaxi162Scf01294g00031	Frameshift (Axs and Exs)	AP2	*ERF110*

For each gene in the final dataset (Gene ID column), the nature of the high-impact mutation, the predicted transcription factor family, and the closest *Arabidopsis* homolog gene symbol are listed.

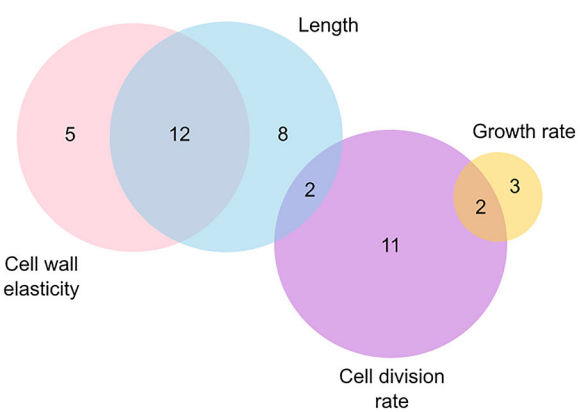

Figure 4. Transcriptome analysis shows greater correlation with cell wall elasticity and style length. Venn plot showing the overlap of the growth-related traits correlated with transcription factors displaying allele-specific expression and/or high-impact polymorphism in their coding sequence. Full details regarding the coefficient values can be found in Table S2.

genomic resources in pollinator-driven speciation models such as *Mimulus*, *Aquilegia,* and *Phlox* as well as our *Petunia* system are still limited. Second, when searching for genes involved in the natural variation of a given trait, the comparison is by definition not between mutant and isogenic wild-type but instead between (nascent) species. Thus, almost any gene in the genome will display sequence polymorphisms, and causative mutations need to be filtered out from a background of divergence caused by genetic drift or associations with unrelated traits. Third, interspecific sequence divergence also causes suppression of recombination, leading to large stretches of DNA in which genetic markers cluster. In fact, linkage can promote the evolution of complex traits by preventing the breakup of favorable gene combinations (Wu and Ting, 2004; Kirkpatrick and Barton, 2006; Schwander *et al.*, 2014). In *Petunia*, five QTLs related to pollinator preference were mapped to such a supergene on chromosome II (Hermann *et al.*, 2013). While this nicely conforms to theory, it makes gene identification by fine mapping a tedious enterprise.

To identify potential speciation genes without *a priori* knowledge of the pathways involved, we devised an approach that combines state-of-the-art transcriptomics with careful characterization of the developmental process as well as publicly available information about comparable developmental processes. Starting from the global transcription factor pool of predicted *Petunia* protein sequences, we applied the workflow shown in Figure 1(b). The amount of candidate genes was strongly reduced in the process, the most stringent steps being ASE-based filtering and functional SNP analysis. As a result, we arrived at a set of 43 candidate genes that were all correlated with at least one of the studied growth traits and were either differentially expressed in cis or had a high-impact SNP in their coding region. Most of the genes identified were preferentially expressed at an early stage and correlated with cell division or were expressed at a later stage and correlated negatively with cell wall elasticity with almost no overlap between these gene sets (Figure 5). The negative correlation with cell wall elasticity suggests that this set of genes is stiffening the cell wall and thus regulating growth arrest. Together, this suggests that style length is regulated independently at two distinct stages and targeting different processes. These data support a view where final style length is a product of early cell division and later cell wall stiffening to trigger growth arrest. The genes that regulate these processes are potential causative agents for the interspecific differences in style morphology.

Prominently absent from our dataset is the homolog of *LO2* (*PRE1* in *Arabidopsis*), an HLH-encoding gene previously shown to be responsible for the style length reduction that occurred during tomato domestication (Chen *et al.*, 2007). However, we identified a candidate HLH protein, BIGPETAL (BPE), known to regulate *Arabidopsis* petal size through its effect on cell expansion (Szécsi *et al.*, 2006). For functional validation, we performed VIGS on six candidate genes. Four of them, including *BPE*, showed no differences. In contrast, inactivation of *EOBI* and *EOBII* had a significant effect on style length, which makes these genes prime candidates for further functional analyses, such as stable overexpression and CRISPR/CAS9. Ultimately, it will be necessary to demonstrate an association between gene expression/gene function and the trait in natural populations.

MYB transcription factors comprise 14 out of the 43 members of the final dataset and are the only enriched

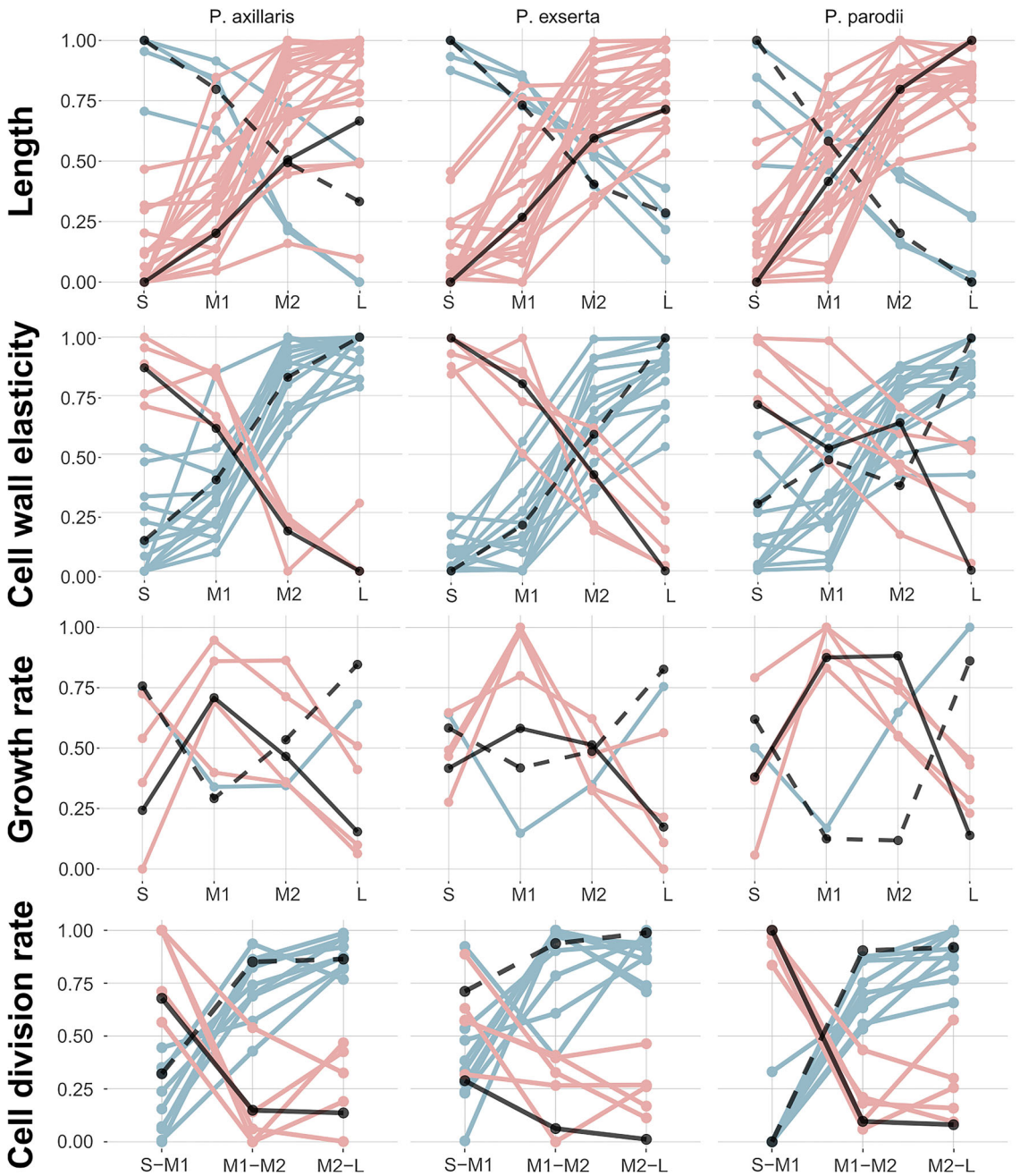

Figure 5. Expression of the majority of genes of interest correlates strongly positively with style length or negatively with cell wall elasticity. Over the growth stages (S, M1, M2, L), feature-scaled values of traits (solid black line) and correlating gene expression (red if positive, blue if negative) are shown. The dashed black line shows an opposite to the trait trend, the trait itself is labeled on the left. For cell division rate, the fold change in cell number was calculated and compared to the log2 fold change of gene expression between adjacent stages of the experiment.

class of genes. This is in line with the thought-provoking suggestion that MYBs may be preferred targets of pollination syndrome evolution (Yuan *et al.*, 2013). MYB factors AN2 and MYB-FL and their homologs in *Mimulus* (and probably many other plant species) are key regulators of the biosynthesis of floral anthocyanins and UV-absorbing flavonols, respectively (Quattrocchio *et al.*, 1999; Yuan *et al.*, 2013; Sheehan *et al.*, 2016). Their action at the late steps in these pathways is thought to reduce pleiotropy and make them hotspots for evolutionary change.

Three of the MYBs identified here, ODO1, EOBI, and PhMYB4, were previously identified as the regulators of floral scent emission (Van Moerkercke *et al.*, 2011; Colquhoun *et al.*, 2011a; Spitzer-Rimon *et al.*, 2012). A further

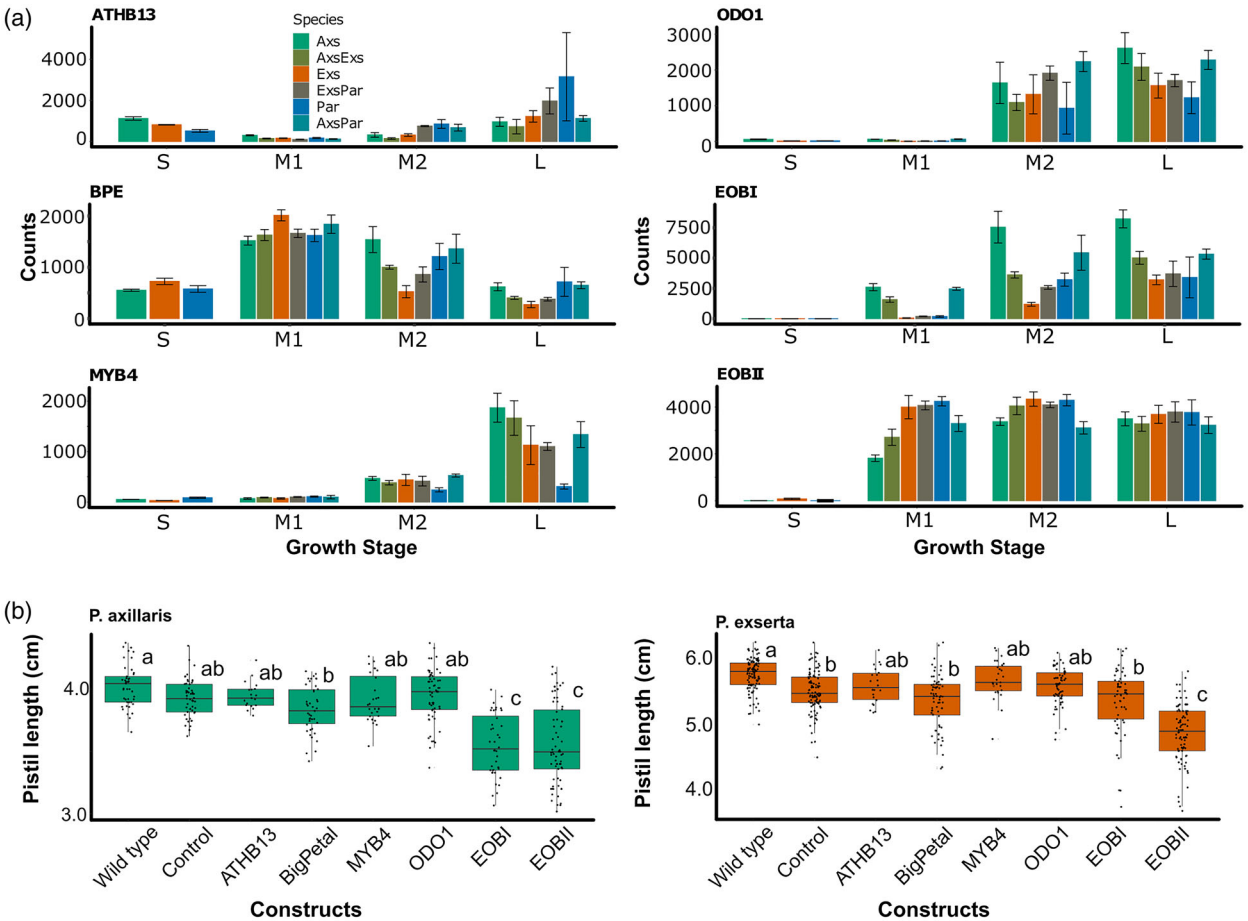

Figure 6. Functional validation of selected candidate genes.
(a) Expression pattern of the candidate genes selected for functional analysis. Barplots represent mean normalized count levels, error bars show mean counts ± SD.
(b) Validation by virus-induced gene silencing (VIGS). Boxplots representing pistil length distribution from *P. axillaris* (left panel) and *P. exserta* plants (right panel) that are not infected, infected with the empty VIGS vector, or infected with the VIGS vector containing a target sequence of each of the candidate genes. Letters on each boxplot represent significance groups as determined by the Tukey HSD test.

scent-related MYB factor, EOBII, is differentially expressed and shows a correlation with growth traits. Unlike the late-acting AN2 and MYB-FL, ODO1, as well as EOBI and EOBII, regulates the shikimate pathway, which leads to the synthesis of the aromatic amino acid phenylalanine (Phe). Phe is the precursor for the entire flavonoid/phenylpropanoid pathway, whose products are not only floral pigments and volatiles but also a plethora of compounds with roles in growth, development, and interactions with the biotic and abiotic environment (Liu *et al.*, 2015). Although no significant difference was observed in plants infected with the *ODO1* VIGS construct (Figure 6b), our results suggest that these genes involved in the regulation of scent in *Petunia* might also be responsible for interspecific variation of style length between wild *Petunia* species. A previous study supports this hypothesis, with EOBII having been shown to inhibit flower opening (Colquhoun *et al.*, 2011b). Speculations that 'one pathway rules them all' might gain credence from further flavonoid/phenylpropanoid pathway analysis.

Identified candidate speciation genes are prime targets for future research. Our results can be used as a starting point for subsequent analyses, ultimately helping to understand the origins of *Petunia* species. Importantly, the employed strategy can be adapted to gene identification for most traits in most non-model organisms. This pipeline also has the potential to contribute to our understanding of the evolution of organ size and shape, a major open question in developmental biology.

EXPERIMENTAL PROCEDURES

Growth conditions and growth rate estimates

Three species of *Petunia* were used: *P. axillaris* (Axs), *P. exserta* (Exs), and *P. parodii* (Par). *Petunia* plants were grown in a greenhouse with 16-h light and daytime temperatures of 22–26°C and

night temperatures of 16–18°C. F_1s were generated for each pair of species, Axs x Exs; Axs x Par; Exs x Par. Bud lengths were measured every few hours and the time of flower opening was recorded for 88 parental buds and 61 F_1 buds (Figure S2). The length is plotted against time until opening. Local polynomial regression fitting was used to predict bud length from time. To relate style length to bud length, 55 buds were dissected and the bud length and style length relationship was defined using local polynomial regression fitting. By combining these two predictions, style length was then expressed in terms of time until flower opening.

Characterizing mechanical properties

The mechanical properties of the styles were characterized using a modified automated micro-extensometer coupled to a light microscope (Robinson *et al.*, 2017). To eliminate the influence of turgor pressure, samples were frozen, then thawed prior to the experiment. The ends of the style were removed and the cell contents and water were removed by squashing the sample. The sample was mounted between two horizontal plates. One of the plates is motorized with a nanopositioner and applies deformation to the sample by varying the distance between the plates. The other plate is passively attached to a force sensor, which reports force acting on the sample. The moving plate performs precisely controlled movements while the sensing plate measures the force. The samples were subjected to cyclic loading at a force of 10 mN for at least three rounds of oscillations. Two dots were added to the style, and images were collected every 2 sec. The ACME tracker was used to automatically follow these two dots in the successive images. The coordinates of the dots were used to compute the strain on the samples. Stars indicate values that are significantly different from the other values of the other species.

Cell division rate estimates

Style tissue from the different stages and species were harvested, flash-frozen in liquid nitrogen, and stored at −80°C. The cells in the long stage styles were lignified and could be observed due to autofluorescence of the lignin without any need for staining. The other styles were cut in half to aid penetration of the stain and stained with calcofluor and KOH buffer or propidium iodide. The samples were left for a few hours until the tissue was completely stained. The method of visualizing the cells was different for the different species and stages as they have different lengths, cross-sections, and wall properties. Sections of style tissue were imaged using confocal microscopy and cell lengths were measured in ImageJ. The length of each style was measured and divided by the average cell length to determine an estimate of the number of cells.

The workflow

1) Identification of transcription factors expressed in the style.

Global transcription factor assay—In order to identify putative transcription factors in the *Petunia* genome, publicly available sequences of *Petunia* proteins (ftp://ftp.solgenomics.net/genomes/Petunia_axillaris/annotation/) were scanned against 57 Pfam (Finn *et al.*, 2013) domain signatures of known plant transcription factors, using the list from (Rushton *et al.*, 2008) as a reference with an E-value of the hit of <0.01 as a filtering threshold. Additionally, as some of the discovered factors do not have a domain signature yet, a simple scan using BLAST+ v2.6.0 (Camacho *et al.*, 2009) was performed for a number of such sequences to identify their homologs. The final transcription factor dataset contains 2284 protein IDs. In order to find the closest homolog in *Arabidopsis*, post-filtering transcription factors were BLAST-scanned against *Arabidopsis* proteins, and then the best hit was back-scanned to the *Petunia* sequence set.

RNA extraction and sequencing—Plants were grown as for the growth measurements. Styles were harvested at 6 p.m., flash-frozen in liquid nitrogen, and then stored at −80°C. RNA was extracted using the Qiagen RNA easy kit according to the manufacturer's instructions. RNA was quantified with a Nanodrop ND-1000 (Thermo Fisher, Waltham, MA, USA) and checked on a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). At least three replicas were performed per stage, per species or hybrid. For the small stage, 20 styles were used, for the other stages at least five styles were pooled per replica. The replicas are independent biological replicas from different plants. Libraries were prepared using Illumina TruSeq PE Cluster Kit v3 and RNA material was sequenced in the Lausanne Genomic Technologies Facility (Lausanne, Switzerland) on an Illumina HiSeq 2500 platform (Illumina, Foster City, CA, USA) in a single-end mode with a read length of 100 bp. All raw sequencing data were submitted as BioProject PRJNA533335 to NCBI.

RNAseq data analysis—Raw read data were quality-checked using FastQC software v. 0.11.2 with a subsequent filtering of the reads based on the final length (at least 60 bp) and the quality score of the bases in the read (at least Q25), along with removing the Illumina adaptor sequences, performed by the fastq-mcf tool, which is a part of the ea-utils package v. 1.1.2 (Aronesty, 2013). Filtered reads were then mapped to the *P. axillaris* reference draft genome v. 1.6.2 using STAR v. 2.5.0b (Dobin *et al.*, 2013) in a basic two-pass mode with a custom edit of the *P. axillaris* annotation (original accessible at https://solgenomics.net/organism/Petunia_axillaris/genome) as a reference for splice junctions with --sjdbOverhang 99 and ignoring reads that map more than 20 times in total. The resulting readmaps along with the same custom annotation served as a base for read counting performed by the htseq-count tool v. 0.6. (Anders *et al.*, 2015). The calculations were performed on UBELIX (http://www.id.unibe.ch/hpc), the HPC cluster at the University of Bern. The count data obtained from htseq-count were processed using the DESeq2 v. 1.12.4 (Love *et al.*, 2014) package in the R v. 3.3.2 computing environment (R Development Core Team, 2016). In order to exclude non- and lowly expressed genes from the analysis, after running the DESeq normalization process the data were filtered with the HTSfilter (v. 1.12.0) (Rau *et al.*, 2013) package (resulting s = 27.9).

Samples were clustered using the R PoiClaClu package v. 1.0.2 (Witten, 2011). During the data preparation step a measure against possible pollen contamination of the long stage samples was included – a group of genes annotated as pollen-specific was found exhibiting a highly similar pattern of expression (average correlation coefficient per gene within the group from 0.95 to 0.98) with counts above filtering threshold present only at the latest stage of the experiment in all studied species. A decision to remove all genes correlating with any of the members of the initial group with a positive coefficient of at least 0.95 was made, excluding 488 genes from the analysis (22 005 genes remaining in the analysis). Counts were normalized using an rlog-transform in DESeq2 and for each gene mean counts were computed over the sample replicates. These averaged and normalized counts were used in the correlation analyses.

2) Correlation of transcript levels with growth-associated traits. The values for various growth and mechanical properties of each experiment stage were correlated against the filtered mean normalized expression data to obtain PCCs. For final pistil length, growth rate, and cell division rate, the expression of genes across all species and hybrids together was correlated with the trait values. The correlation was performed for each trait individually but using all of the available gene expression for all of the stages, species, and hybrids together. For cell wall elasticity, only gene expression of the parental species was used for coefficient calculation. To compare the general linear correlation tendency of transcripts and traits, a permutation test was performed as described in (Baute *et al.*, 2016), but with analyzing the top 5% by the absolute value of correlation instead of q0.01 and q0.99 for the purpose of this analysis. Resulting permutation mean values were significantly lower than the threshold of 0.75 chosen for this analysis, with values being: for style length 0.341, for style growth rate 0.346, for cell division rate 0.408, and for cell wall elasticity 0.466.

3) Allele-specific expression (ASE) in interspecific F_1s. Allelic coverage for variant positions in *P. axillaris* x *P. exserta* F_1 and *P. axillaris* x *P. parodii* F_1 (Axs x Exs F_1 and Axs x Par F_1) read maps was detected with ASEReadCounter implemented in GATK (DePristo *et al.*, 2011; Castel *et al.*, 2015). Variant positions were filtered using hard thresholds with variants having any of the DP < 10, QD < 2, MQ < 40, FS > 60, MQRankSum < −12.5, and ReadPosRankSum < −8.0 removed from further analysis. Clustered SNPs with more than three occurrences in a window of five were also omitted.

Analysis of allelic imbalance was conducted in R software with the package MBASED v.1.2.0 (Mayba *et al.*, 2014) using rounded read counts averaged over the biological replicates and one-sample mode, which identifies within-sample deviations using a null-hypothesis of equal allele expression. A correction for potential allele bias was implemented with the help of the package using the single-SNP genes with more than 30 reads detected for both alleles and low allelic disequilibrium (less than 1.4-fold difference in read count, final rho estimated as 0.00308). The results of the assay were controlled for the false discovery rate by applying the Benjamini–Hochberg corrections procedure (Benjamini *et al.*, 2001). The R script used for RNAseq data processing, the ASE data, and the gene annotations can be found in Data S2–S5.

4) Analysis of functionally relevant SNPs in the coding region. In order to identify functional SNPs, variant maps obtained for ASE analysis were scanned using SNPeff v.4.3T (Cingolani *et al.*, 2012). Detected mutations from high-impact category (e.g., frameshifts, loss and gain of stop/start codon, rare amino acid variants, and exon splice site mutations) were considered for further analysis. In order to minimize the number of false-positive hits, a filtering method similar to that described in (Xu *et al.*, 2019) was used on the results. To further validate the existence of target genes and SNPs, comparison of genomic sequences of genes of interest from different species was performed using BLAST+. A draft genome needed to be assembled for *P. parodii de novo*, as none was available.

DNA extraction, sequencing, and assembly of P. parodii—*Petunia parodii* S7 was kindly provided by R. Koes, University of Amsterdam, and maintained locally by selfing. Plants were grown axenically in tissue culture containers and used for DNA extraction as described in (Bombarely *et al.*, 2016). DNA material was sequenced on an Illumina Hiseq3000 platform, producing 150 bp paired-end reads with 400 bp library insert size. Reads were quality-controlled using trimmomatic v. 0.33 (Bolger *et al.*, 2014); the results were checked using FastQC software v. 0.11.2 (Leggett *et al.*, 2013). For the assembly, an optimal kmer value of 47 was estimated with the help of jellyfish v. 2.1.0 (Zimin *et al.*, 2013) with the -m 21 parameter. Assembly of quality-controlled reads was performed by SOAPdenovo2 v. 2.04.240 (Luo *et al.*, 2012) with parameters max_rd_len = 150, avg_ins = 400, reverse_seq = 0, rd_len_cutoff = 150, rank = 1, pair_num_cutoff = 4, and map_len = 32. A summary of the assembled draft genome can be seen in Table 2.

In order to estimate the gene content in the assembly, *P. axillaris* full cDNA was aligned to the draft genome using GMAP (version of June 20, 2017) (Wu and Watanabe, 2005) with default parameters, resulting in the detectable presence of approximately 30 000 genes in the assembly. Of these genes, 12,827 were present in the genome with a length of at least 95% of the length of the query sequence. All of the bash scripts used for data transformation and analysis can be found in Data S1.

Virus-induced gene silencing (VIGS)

We first selected coding region fragments in the targeted genes that are the most specific to avoid off-targets. These fragments were amplified by PCR from cDNA using forward and reverse primers containing, respectively, BamHI and EcoRI restriction sites for subsequent cloning in the pTRV2-MCS plasmid (ABRC code: CD3-1040) and introduction into *Agrobacterium tumefaciens* strain GV3101 (Table S5). Then, VIGS experiments were performed as described previously (Spitzer-Rimon *et al.*, 2013). Briefly, pTRV1 (ABRC code: CD3-1039) and each pTRV2 derivative were prepared and mixed in a 1:1 ratio prior to infection of *Petunia* plants (approximately 6-leave stage). The infection was carried out by removing the shoot apex and applying a drop of inoculum to the cut surface of the stem. The two branches closest to the infection points were used for the pistil phenotyping in 2 days post-anthesis flowers. We used two different controls, that is, buffer control and plants infected with the empty vector (pTRV2-MCS), and we observed that the latter control group displayed in general smaller pistil size (Figure 6). Although the plants looked overall healthy, we attributed this difference to a viral response since previous studies reported that infection with the empty virus produced severe viral symptoms such as lesions, stunting, and death in several Solanaceae species; see for example Hartl *et al.*, 2008; Wu *et al.*, 2011; Broderick and Jones, 2014. *P. parodii* was not used for the VIGS experiments as this species is late-flowering and highly susceptible to the empty vector.

DATA STATEMENT

The large-scale datasets including transcriptomes, genomes are freely available at NCBI and the Sol Genomics Network (solgenomics.net). The *P. parodii* data is housed under NCBI bioproject PRJNA533335 with the following accession numbers: JACAFL000000000 (assembly), SRX5714874 (nucleotide reads) and SRX5711291-SRX5711362 (RNASeq reads). The *Petunia axillaris* v4 annotations are available at: https://zenodo.org/record/3922967.

ACKNOWLEDGMENTS

We thank Holly Summers and Hagen Reinhardt for help with sample collection, RNA extraction, and phenotyping, and Christopher Ball and Jasmin Sekulovski for expert care of the plants. Korinna

Esfeld and Gina Cannarozzi carefully read and commented on the manuscript. This work was supported by SystemsX.ch, ERC Advanced Grant 741354, Swiss NSF grant 31003A_182340, and University of Bern core funding. Calculations were performed on UBELIX (http://www.id.unibe.ch/hpc), the HPC cluster at the University of Bern.

AUTHOR CONTRIBUTIONS

TY performed the bioinformatics analyses; SR and VT performed the biomechanical assays; MH and SR performed the phenotyping and the RNAseq experiment; MH performed the VIGS experiments; CK supervised the project. All the authors participated in the writing of the manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Heatmap generated based on Poisson's distance between studied samples.

Figure S2. Growth curves for F_1s.

Table S1. Measured growth-related traits.

Table S2. Putative candidate transcription factors.

Table S3. Expression attributes.

Table S4. Enrichment assay data.

Table S5. Sequences of primers used.

Data S1. Supporting information.

Data S2. RNAseq_v3R.

Data S3. Peaxi162 annotation_v4gff.

Data S4. AxEx_CDS_style.csv (1.0).

Data S5. AxPa_CDS_style.csv (1.0).

REFERENCES

Alexandersson, R. and Johnson, S.D. (2002) Pollinator-mediated selection on flower-tube length in a hawkmoth-pollinated Gladiolus (Iridaceae). *Proc. R Soc. B Biol. Sci.* **269**, 631–636.

Amrad, A., Moser, M., Mandel, T., de Vries, M., Schuurink, R.C., Freitas, L. and Kuhlemeier, C. (2016) Gain and loss of floral scent production through changes in structural genes during pollinator-mediated speciation. *Curr. Biol.* **26**, 3303–3312.

Anders, S., Pyl, P.T. and Huber, W. (2015) HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.

Armbruster, W.S. and Muchhala, N. (2009) Associations between floral specialization and species diversity: cause, effect, or correlation? *Evol. Ecol.* **23**, 159–179.

Aronesty, E. (2013) Comparison of sequencing utility programs. *Open Bioinformatics J.* **7**, 1–8.

Battat, M., Eitan, A., Rogachev, I., Hanhineva, K., Fernie, A., Tohge, T., Beekwilder, J. and Aharoni, A. (2019) A MYB triad controls primary and phenylpropanoid metabolites for pollen coat patterning. *Plant Physiol.* **180**, 87–108.

Baute, J., Herman, D., Coppens, F., De Block, J., Slabbinck, B., Dell'Aqcua, M., Pè, M.E., Maere, S., Nelissen, H. and Inzé, D. (2016) Combined large-scale phenotyping and transcriptomics in maize reveals a robust growth regulatory network. *Plant Physiol.* **170**, 1848–1867.

Beemster, G.T.S. and Baskin, T.I. (1998) Analysis of cell division and elongation underlying the developmental acceleration of root growth in Arabidopsis thaliana. *Plant Physiol.* **116**, 1515–1526.

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N. and Golani, I. (2001) Controlling the false discovery rate in behavior genetics research. *Behav. Brain Res.* **125**, 279–284.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Bombarely, A., Moser, M., Amrad, A. *et al.* (2016) Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida. *Nat. Plants*, **2**, 16074.

Braybrook, S.A. and Jönsson, H. (2016) Shifting foundations: the mechanical cell wall and development. *Curr. Opin. Plant Biol.* **29**, 115–120.

Broderick, S.R. and Jones, M.L. (2014) An optimized protocol to increase virus-induced gene silencing efficiency and minimize viral symptoms in Petunia. *Plant Mol. Biol. Rep.* **32**, 219–233.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.

Castel, S.E., Levy-Moonshine, A., Mohammadi, P., Banks, E. and Lappalainen, T. (2015) Tools and best practices for data processing in allelic expression analysis. *Genome Biol.* **16**, 195.

Chen, K.Y., Cong, B., Wing, R., Vrebalov, J. and Tanksley, S.D. (2007) Changes in regulation of a transcription factor lead to autogamy in cultivated tomatoes. *Science*, **318**, 643–645.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. and Ruden, D.M. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, **6**, 80–92.

Colquhoun, T.A., Kim, J.Y., Wedde, A.E., Levin, L.A., Schmitt, K.C., Schuurink, R.C. and Clark, D.G. (2011a) PhMYB4 fine-tunes the floral volatile signature of Petunia x hybrida through PhC4H. *J. Exp. Bot.* **62**, 1133–1143.

Colquhoun, T.A., Schwieterman, M.L., Wedde, A.E. *et al.* (2011b) EOBII controls flower opening by functioning as a general transcriptomic switch. *Plant Physiol.* **156**, 974–984.

Cronk, Q. and Ojeda, I. (2008) Bird-pollinated flowers in an evolutionary and molecular context. *J. Exp. Bot.* **59**, 715–727.

Czesnick, H. and Lenhard, M. (2015) Size control in plants—lessons from leaves and flowers. *Cold Spring Harb. Perspect. Biol.* **7**, 1–16.

DePristo, M.A., Banks, E., Poplin, R. *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

Doebley, J. and Lukens, L. (1998) Transcriptional regulators and the evolution of plant form. *Plant Cell*, **10**, 1075–1082.

Fenster, C.B., Armbruster, W.S., Wilson, P., Dudash, M.R. and Thomson, J.D. (2004) Pollination syndromes and floral specialization. *Annu. Rev. Ecol. Evol. Syst.* **35**, 375–403.

Finn, R.D., Bateman, A., Clements, J. *et al.* (2013) Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–D230.

Hartl, M., Merker, H., Schmidt, D.D. and Baldwin, I.T. (2008) Optimized virus-induced gene silencing in *Solanum nigrum* reveals the defensive function of leucine aminopeptidase against herbivores and the shortcomings of empty vector controls. *New Phytol.* **179**, 356–365.

Hermann, K., Klahre, U., Venail, J., Brandenburg, A. and Kuhlemeier, C. (2015) The genetics of reproductive organ morphology in two Petunia species with contrasting pollination syndromes. *Planta*, **241**, 1241–1254.

Hermann, K., Klahre, U., Moser, M., Sheehan, H., Mandel, T. and Kuhlemeier, C. (2013) Tight genetic linkage of prezygotic barrier loci creates a multifunctional speciation island in Petunia. *Curr. Biol.* **23**, 873–877.

Hoballah, M.E., Gubitz, T., Stuurman, J., Broger, L., Barone, M., Mandel, T., Dell'Olivo, A., Arnold, M. and Kuhlemeier, C. (2007) Single gene-mediated shift in pollinator attraction in Petunia. *Plant Cell*, **19**, 779–790.

Jia, Y., Jia, M.H., Wang, X. and Liu, G. (2012) Indica and Japonica crosses resulting in linkage block and recombination suppression on rice chromosome 12. *PLoS One*, **7**, e43066.

Kirkpatrick, M. and Barton, N. (2006) Chromosome inversions, local adaptation and speciation. *Genetics*, **173**, 419–434.

Leggett, R.M., Ramirez-Gonzalez, R.H., Clavijo, B.J., Waite, D. and Davey, R.P. (2013) Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Front. Genet.* **4**, 288.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Subgroup, G.P.D.P. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Liu, G. and Thornburg, R.W. (2012) Knockdown of MYB305 disrupts nectary starch metabolism and floral nectar production. *Plant J.* **70**, 377–388.

Liu, J.Y., Osbourn, A. and Ma, P.D. (2015) MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *Mol. Plant*, **8**, 689–708.

Lockhart, J.A. (1965) An analysis of irreversible plant cell elongation. *J. Theor. Biol.* **8**, 115–120.

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.

Luo, R., Liu, B., Xie, Y. *et al.* (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*, **1**, 18.

Mayba, O., Gilbert, H.N., Liu, J., Haverty, P.M., Jhunjhunwala, S., Jiang, Z., Watanabe, C. and Zhang, Z. (2014) MBASED: allele-specific expression detection in cancer tissues and cell lines. *Genome Biol.* **15**, 405.

Ostberg, C.O., Hauser, L., Pritchard, V.L., Garza, J.C. and Naish, K.A. (2013) Chromosome rearrangements, recombination suppression, and limited segregation distortion in hybrids between Yellowstone cutthroat trout (*Oncorhynchus clarkii bouvieri*) and rainbow trout (*O. mykiss*). *BMC Genom.* **14**, 570.

Pastinen, T. (2010) Genome-wide allele-specific analysis: insights into regulatory variation. *Nat. Rev. Genet.* **11**, 533–538.

Peaucelle, A., Wightman, R. and Höfte, H. (2015) The control of growth symmetry breaking in the Arabidopsis Hypocotyl. *Curr. Biol.* **25**, 1746–1752.

Peng, F., Byers, K.J.R.P. and Bradshaw, H.D. Jr (2017) Less is more: independent loss-of-function OCIMENE SYNTHASE alleles parallel pollination syndrome diversification in monkeyflowers (Mimulus). *Am. J. Bot.* **104**, 1055–1059.

Quattrocchio, F., Wing, J., van der Woude, K., Souer, E., de Vetten, N., Mol, J. and Koes, R. (1999) Molecular analysis of the anthocyanin2 gene of petunia and its role in the evolution of flower color. *Plant Cell*, **11**, 1433–1444.

Raguso, R.A. and Willis, M.A. (2002) Synergy between visual and olfactory cues in nectar feeding by naive hawkmoths, *Manduca sexta*. *Anim. Behav.* **64**, 685–695.

Rau, A., Gallopin, M., Celeux, G. and Jaffrézic, F. (2013) Data-based filtering for replicated high-throughput transcriptome sequencing experiments. *Bioinformatics*, **29**, 2146–2152.

Ren, X., Li, R., Wei, X. *et al.* (2018) Genomic basis of recombination suppression in the hybrid between *Caenorhabditis briggsae* and *C. nigoni*. *Nucleic Acids Res.* **46**, 1295–1307.

Ribone, P.A., Capella, M. and Chan, R.L. (2015) Functional characterization of the homeodomain leucine zipper I transcription factor AtHB13 reveals a crucial role in Arabidopsis development. *J. Exp. Bot.* **66**, 5929–5943.

Robinson, S., Huflejt, M., Barbier de Reuille, P., Braybrook, S.A., Schorderet, M., Reinhardt, D. and Kuhlemeier, C. (2017) An automated confocal micro-extensometer enables in vivo quantification of mechanical properties with cellular resolution. *Plant Cell*, **29**, 2959–2973.

Rushton, P.J., Bokowiec, M.T., Han, S., Zhang, H., Brannock, J.F., Chen, X., Laudeman, T.W. and Timko, M.P. (2008) Tobacco transcription factors: novel insights into transcriptional regulation in the Solanaceae. *Plant Physiol.* **147**, 280–295.

Sapir, N. and Dudley, R. (2013) Implications of floral orientation for flight kinematics and metabolic expenditure of hover-feeding hummingbirds. *Funct. Ecol.* **27**, 227–235.

Schwander, T., Libbrecht, R. and Keller, L. (2014) Supergenes and complex phenotypes. *Curr. Biol.* **24**, R288–R294.

Sheehan, H., Moser, M., Klahre, U., Esfeld, K., Dell'Olivo, A., Mandel, T., Metzger, S., Vandenbussche, M., Freitas, L. and Kuhlemeier, C. (2016) MYB-FL controls gain and loss of floral UV absorbance, a key trait affecting pollinator preference and reproductive isolation. *Nat. Genet.* **48**, 159–166.

Spitzer-Rimon, B., Cna'ani, A. and Vainstein, A. (2013) Virus-Aided gene expression and silencing using TRV for functional analysis of floral scent-related genes. *Methods Mol. Biol.* **975**, 139–148.

Spitzer-Rimon, B., Marhevka, E., Barkai, O., Marton, I., Edelbaum, O., Masci, T., Prathapani, N.-K., Shklarman, E., Ovadis, M. and Vainstein, A. (2010) EOBII, a gene encoding a flower-specific regulator of phenylpropanoid volatiles' biosynthesis in petunia. *Plant Cell*, **22**, 1961–1976.

Spitzer-Rimon, B., Farhi, M., Albo, B. *et al.* (2012) The R2R3-MYB-like regulatory factor EOBI, acting downstream of EOBII, regulates scent production by activating ODO1 and structural scent-related genes in petunia. *Plant Cell*, **24**, 5089–5105.

Stebbins, G.L. (1970) Adaptive radiation of reproductive characteristics in angiosperms, I: pollination mechanisms. *Annu. Rev. Ecol. Syst.* **1**, 307–326.

Szécsi, J., Joly, C., Bordji, K., Varaud, E., Cock, J.M., Dumas, C. and Bendahmane, M. (2006) BIGPETALp, a bHLH transcription factor is involved in the control of Arabidopsis petal size. *EMBO J.* **25**, 3912–3920.

Thomson, J.D. and Wilson, P. (2008) Explaining evolutionary shifts between bee and hummingbird pollination: convergence, divergence, and directionality. *Int. J. Plant Sci.* **169**, 23–38.

Van Moerkercke, A., Haring, M.A. and Schuurink, R.C. (2011) The transcription factor EMISSION OF BENZENOIDS II activates the MYB ODORANT1 promoter at a MYB binding site specific for fragrant petunias. *Plant J.* **67**, 917–928.

Venail, J., Dell'Olivo, A. and Kuhlemeier, C. (2010) Speciation genes in the genus Petunia. *Philos. Trans. R Soc. Lond. Series B Biol. Sci.* **365**, 461–468.

Verdonk, J.C., Haring, M.A., van Tunen, A.J. and Schuurink, R.C. (2005) ODORANT1 regulates fragrance biosynthesis in petunia flowers. *Plant Cell*, **17**, 1612–1624.

Wessinger, C.A. and Rausher, M.D. (2012) Lessons from flower colour evolution on targets of selection. *J. Exp. Bot.* **63**, 5741–5749.

Whitney, H.M., Chittka, L., Bruce, T.J.A. and Glover, B.J. (2009) Conical epidermal cells allow bees to grip flowers and increase foraging efficiency. *Curr. Biol.* **19**, 948–953.

Witten, D.M. (2011) Classification and clustering of sequencing data using a Poisson model. *Ann. Appl. Stat.* **5**, 2493–2518.

Wittkopp, P.J. and Kalay, G. (2012) Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* **13**, 59–69.

Wu, C.-I. and Ting, C.-T. (2004) Genes and speciation. *Nat. Rev. Genet.* **5**, 114–122.

Wu, C.J., Jia, L.L. and Goggin, F. (2011) The reliability of virus-induced gene silencing experiments using tobacco rattle virus in tomato is influenced by the size of the vector control. *Mol. Plant Pathol.* **12**, 299–305.

Wu, T.D. and Watanabe, C.K. (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, **21**, 1859–1875.

Xu, Y.-C., Niu, X.-M., Li, X.-X., He, W., Chen, J.-F., Zou, Y.-P., Wu, Q., Zhang, Y.E., Busch, W. and Guo, Y.-L. (2019) Adaptation and phenotypic diversification in arabidopsis through loss-of-function mutations in protein-coding genes. *Plant Cell*, **31**, 1012–1025.

Yuan, Y.-W., Byers, K.J.R.P., Bradshaw, H.D. Jr. (2013) The genetic control of flower-pollinator specificity. *Curr. Opin. Plant Biol.* **16**, 422–428.

Zimin, A.V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S.L. and Yorke, J.A. (2013) The MaSuRCA genome assembler. *Bioinformatics*, **29**, 2669–2677.